U4 Estimadores puntuales e intervalos de confianza

Las VAs, como ya es sabido, requieren de un conjunto de constantes para poder trabajar con ellas. Sin embargo, si bien venimos trabajando con estas constantes como dato, en la vida real no es tan simple. Por ello, se deben estimar a partir de los datos recolectados en las muestras.

PARÁMETROS	ESTIMADORES PUNTUALES
μ MEDIA POBLACIONAL	$\bar{x}$ MEDIA MUESTRAL
σ^2 VARIANZA POBLACIONAL	s^2 VARIANZA MUESTRAL
p PROPORCIÓN POBLACIONAL	$\hat{p}$ PROPORCIÓN MUESTRAL

- MEDIA MUESTRAL

Conjunto de VAs independientes idénticamente distribuidas
VAIID de ahora en adelante

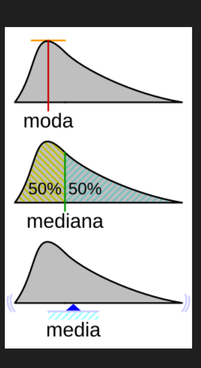
- independientes → Una VA no influye en la ocurrencia de la otra
- idénticamente distribuidas → Todas tienen la misma distribución, media y varianza

Se define:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

- Características
  - Mediana → x donde se encuentra el percentil 50
    - Si n es par:

$$Med = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$$



Ej:

x_1	x_2	x_3	x_4
3	6	8	1

$n=4$

ORDENAR DE MEN. A MAY.

x_1	x_2	x_3	x_4
1	3	6	8

$$Med = \frac{x_{\frac{4}{2}} + x_{\frac{4}{2}+1}}{2} = \frac{x_2 + x_3}{2} = \frac{3+6}{2} = 4,5$$

→ Puede llegar a ser un número con decimales

- Si n es impar:

$$Med = x_{\frac{n+1}{2}}$$

Ej:

x_1	x_2	x_3	x_4	x_5
3	6	8	1	9

$n=5$

ORDENAR DE MEN. A MAY.

x_1	x_2	x_3	x_4	x_5
1	3	6	8	9

$$Med = x_{\frac{n+1}{2}} = x_{\frac{5+1}{2}} = x_{\frac{6}{2}} = x_3 = 6$$

- Moda → valor más repetido en el set de datos

Ej:

x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	1	2	5	7	7	7

2 veces (x_1, x_2); 1 vez (x_3); 1 vez (x_4); 3 veces (x_5, x_6, x_7) MODA

$Mo = x_5$, x_6, x_7 } Puedo dar cualquiera de estas como solución.

- SESGO $B(\hat{\theta})$

Un estimador es sesgado cuando la esperanza de dicho $\hat{\theta}$ NO es igual al parámetro estimado. Son insesgados $\bar{x}$, $\hat{p}$ y s^2

SESGO: $E(\hat{\theta}) - \text{parámetro} = a$

↳ si $a = 0 \Rightarrow \hat{\theta}$ es insesgado

- EFICIENCIA RELATIVA

Sean $\hat{\theta}_1$ y $\hat{\theta}_2$ dos estimadores insesgados:

si $V(\hat{\theta}_1) < V(\hat{\theta}_2) \Rightarrow \hat{\theta}_1$ es más eficiente que $\hat{\theta}_2$

¿Cómo elijo mi estimador?

1. Descarto $\hat{\theta}$ sesgados
2. Comparo varianzas de los θ seleccionados
3. Me quedo con el de menor varianza

- CONSISTENCIA

Un estimador insesgado es consistente si:

$$\lim_{n \to \infty} V(\hat{\theta}) = 0$$

- SUFICIENCIA

Un estimador es suficiente cuando utiliza toda la información muestral

- ERROR CUADRÁTICO MEDIO ECM

$$ECM(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = V(\hat{\theta}) + B(\hat{\theta})$$

↳ si el estimador es insesgado $\Rightarrow B(\hat{\theta}) = 0$

- Teorema central del límite TCL

El TCL estipula que, dada una n-ésima suma de m VAs con esperanza y varianza finitas, la función de distribución de dicha suma se asemeja a una distribución normal.

Si $X_1, X_2, X_3, \cdots, X_n$ son muestras aleatorias extraídas de una población con media μ y varianza σ^2, y si $\bar{x}_n$ es la media muestral de las primeras n muestras, entonces la forma límite de la distribución $Z = \lim_{n \to \infty} \frac{\bar{x}_n - \mu}{\sigma/\sqrt{n}}$ es una normal estándar.

- Esto vale cuando $n \geq 30$
- $\bar{x}$ se comporta de manera similar a μ (aprox. igual o igual)

- Nuevas distribuciones empleadas en esta unidad
  - CHI CUADRADO χ^2_k

Siendo $Z_1, Z_2, Z_3, \cdots, Z_k$ todas VAIID con $Z_k \sim N(0;1)$ y $x = \sum_{i=1}^{n} Z_i$ se dice que $x \sim \chi^2_k$, donde k representa los grados de libertad de la VA.

$$\chi^2_k \sim \frac{(n-1)s^2}{\sigma^2}, \quad k = n-1$$

- $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$
- $E(x) = k$
- $V(x) = 2k$

  - T de student

Se define como:

$$\frac{Z}{\sqrt{\frac{\chi^2_{n-1}}{n-1}}} = \frac{\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{\frac{(n-1)s^2}{\sigma^2}}{n-1}}} = \frac{\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{s^2}{\sigma^2}}} = \frac{\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}}{\frac{s}{\sigma}}$$

$$t_{n-1} \sim \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$$

- INTERVALO DE CONFIANZA

Se emplean para estimar un intervalo al cual pertenece un determinado parámetro. Tiene una cierta probabilidad de ser preciso (o no).
Este intervalo se caracteriza así:

$$P(\hat{\theta}_{inf} \leq \hat{\theta} \leq \hat{\theta}_{sup}) = 1 - \alpha$$

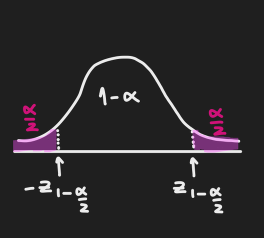
Donde:
- $1-\alpha$ → nivel de confianza
- $\hat{\theta}_{inf}$ → límite inferior
- $\hat{\theta}_{sup}$ → límite superior

Según el estimador solicitado y los datos que se tienen a mano se emplea cada una de las siguientes distribuciones:

DISTRIBUCIÓN NORMALES

Para σ conocido

$$P\left(-Z_{1-\frac{\alpha}{2}} \leq \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq Z_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$



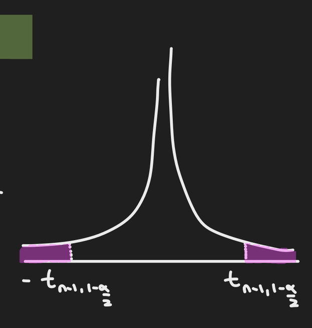
$$-Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} + \bar{x} \leq \mu \leq Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} + \bar{x}$$

$$IC_{1-\alpha} = \left(-Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} + \bar{x} \,;\, Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} + \bar{x}\right)$$

DISTRIBUCIÓN T STUDENT

Para μ. σ desconocido

$$P\left(-t_{n-1,1-\frac{\alpha}{2}} \leq \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} \leq t_{n-1,1-\frac{\alpha}{2}}\right) = 1-\alpha$$



$$-t_{n-1,1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} + \bar{x} \leq \mu \leq t_{n-1,1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} + \bar{x}$$

$$IC_{1-\alpha} = \left(-t_{n-1,1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} + \bar{x} \,;\, t_{n-1,1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} + \bar{x}\right)$$

DISTRIBUCIÓN CHI CUADRADO

Para σ^2

$$P\left(\chi^2_{n-1,\frac{\alpha}{2}} \leq \frac{(n-1)s^2}{\sigma^2} \leq \chi^2_{n-1,1-\frac{\alpha}{2}}\right) = 1-\alpha$$

$$IC_{1-\alpha} = \left(\frac{(n-1)s^2}{\chi^2_{n-1,\frac{\alpha}{2}}} \,.\, \frac{(n-1)s^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}\right)$$

DISTRIBUCIÓN BERNOULLI

Para p

$$\hat{p} = \frac{\sum_{i=1}^{n} X_i \sim Be(p)}{n}$$

$$IC_{1-\alpha} = \left(-Z_{1-\frac{\alpha}{2}} \sqrt{\frac{p \cdot q}{n}} + \hat{p} \,;\, Z_{1-\frac{\alpha}{2}} \sqrt{\frac{p \cdot q}{n}} + \hat{p}\right)$$

($p \cdot q$ → σ)